J Comput Electron (2016) 15:1130–1147
DOI 10.1007/s10825-016-0851-0

Impact of phonon scattering in Si/GaAs/InGaAs nanowires and FinFets: a NEGF perspective

Antonio Martinez[1] · **Anna Price**[1] · **Raul Valin**[1] · **Manuel Aldegunde**[2] · **John Barker**[3]

Published online: 21 July 2016

Abstract This paper reviews our previous theoretical studies and gives further insight into phonon scattering in 3D small nanotransistors using non-equilibrium Green function methodology. The focus is on very small gate-all-around nanowires with Si, GaAs or InGaAs cores. We have calculated phonon-limited mobility and transfer characteristics for a variety of cross-sections at low and high drain bias. The nanowire cross-sectional area is shown to have a significant impact on the phonon-limited mobility and on the current reduction. In a study of narrow Si nanowires we have examined the spatially resolved power dissipation and the validity of Joule's law. Our results show that only a fraction of the power is dissipated inside the drain region even for a relatively large simulated length extension (approximately 30 nm). When considering large source regions in the simulation domain, at low gate bias, a slight cooling of the source is observed. We have also studied the impact of the real part of phonon scattering self-energy on a narrow nanowire transistor. This real part is usually neglected in nanotransistor simulation, whereas we compute its impact on current–voltage characteristic and mobility. At low gate bias, the imaginary part strongly underestimated the current and the mobility by 50 %. At high gate bias, the two mobilities are similar and the effect on the current is negligible.

Keywords Silicon and III–V nanowire field effect transistors · Non-equilibrium Green's functions · Electron–phonon scattering self-energies · Phonon-limited mobility · Local power dissipation

✉ Antonio Martinez
a.e.martinez@swansea.ac.uk

Anna Price
625036@swansea.ac.uk

Raul Valin
raul@raulvalin.es

Manuel Aldegunde
M.A.Aldegunde-Rodriguez@warwick.ac.uk

John Barker
John.Barker@glasgow.ac.uk

[1] College of Engineering, Swansea University, Engineering East, Fabian Way, Crymlyn Burroughs, Swansea SA1 8EN, UK

[2] School of Engineering, University of Warwick, Coventry CV4 7AL, UK

[3] College of Science and Engineering, University of Glasgow, Rankine Building, Oakfield Avenue, Glasgow G12 8LT, UK

1 Introduction

Three-dimensional field effect transistor structures, such as nanowires and FinFETs, have been extensively investigated. These structures provide a protection against short channel effects and excellent electrostatic integrity [1–4]. [1] Their theoretical calculated subthreshold slope is almost 60 mV/dec for a well-balanced structure. Indeed, FinFET structures of approximately 40-nm channel length have been in production for several years. Understanding the miniaturization potential of these devices is of both economic and research importance.

Semiclassical simulation transport methodologies, such as drift diffusion, lose their predictability in the nanometer realm due to the wave nature of the electron. However, the charge integrity of MOSFET transistors seems to brush out some quantum effects. This happens when the effect of tunneling and confinement is negligible. Hence quasi-classical methodologies are still applicable at nanometer

[1] See http://www.itrs.net/for "International Technology Roadmap for Semiconductors."

dimensions. Of course continuous calibration for each particular dimension is still required in order to accurately predict current–voltage characteristics. However, for transistors around tens of nanometers the quantum effect becomes so severe that solely quantum mechanical simulations can provide a reliable description of transport. For silicon devices with cross-sections under $16\,\mathrm{nm}^2$, confinement becomes important and therefore modeling the transport through subbands is strongly recommended. It appears that small gate length (under 10 nm) devices should be ballistic and of course this is true if the cross-section is much larger than $16\,\mathrm{nm}^2$ [5], but when the cross-section becomes smaller, the effective phonon coupling increases inversely proportional to the cross-section [6]. The physical reason is the strong localization of the electron wave function of the channel cross-section. This is reflected through the form factors [5,7] that appear in the calculation of the electron–phonon scattering self-energies in the mode description. They depend on the inverse of the area of the cross-section and are the main reason for this decrease of phonon-limited mobility [6]. Therefore, narrow wires should have decreased phonon-limited mobility. This fact has been found out using a variety of theoretical models [8,9] and has also been confirmed experimentally [10]. Of course, for small device cross-sections, interface scattering plays a very important role, as the volume/area ratio is small. However, here we will concentrate on the phonon scattering mechanism. Electron–electron scattering, surface roughness and photon emission have not been considered. However, their overall effect is negligible because they are elastic processes, which do not change the energy relaxation.

It is customary to only consider the imaginary part of the self-energy in device simulation. Only a handful of authors consider full self-energy [11–14] as the calculation of the real part (or the renormalization of the levels due to electron phonon scattering) needs the computation of the Hilbert transform [15], which is very time-consuming. In addition, the conservation of the density of states requires the real as well as the imaginary part of the self-energy [17]. This is a consequence of the Kramers–Kronig relationship [16] or causality conditions (details in the "Appendix"). The neglect of the real part produces an underestimate of the electron density in the channel of the transistor; this fact has been pointed out by A. Svizhenko [12]. Similar effects have been observed by other authors [11,13]. Ignoring the real part produces an underestimate of mobility at low inversion charges [14].

Another important issue is power dissipation in ultra-high-scale 3D integration [18–20]. Miniaturization compresses a huge number of devices into a small region and, when combined with the heat partially trapped inside the channel of a 3D MOSFET, makes heat release a difficult enterprise. Heat becomes trapped because an oxide material, which has a very low thermal conductivity, surrounds the channel. At device level, it is important to know where heat is dissipated, i.e., drain–channel interface or deep inside the drain. For narrow devices this is a complex issue as it depends on particular inelastic phonon energies and coupling constants and on the number of states available locally or the empty local density of states, which is broadened due to scattering and is dependent on the local potential. All this occurs under highly non-equilibrium conditions for shorter channel lengths. In order to capture these phenomena a non-equilibrium quantum description of transport is required. Datta et al. [21,22] use a NEGF formalism in order to calculate the local power dissipation in a double-barrier resonant tunneling device. In those papers a good description of energy transfer and conservation laws in the NEGF formalism is presented. Mahan's report [23] goes deeper into the conservation laws. Recently [24], we have calculated the local power using the NEGF formalism for Si GAA (gate-all-around) nanowire transistors. The calculation of local power is quite sensitive, as one needs to subtract two large terms in order to calculate the power dissipation (see Sect. 3). This means that the calculation of the local current spectra needs to be done quite accurately to ensure that the local current conservation is fulfilled with a relative error in the current of $10^{-3} - 10^{-4}$. This implies that many Born iterations are required making the calculation computationally expensive.

The improvement in the quality of III–V materials and their interfaces make the fabrication of small III–V transistors a real possibility [25]. The fact that these materials have direct band gap and high mobility makes them attractive for optoelectronic and digital applications. In this work we investigate ultra-scaled nanowire GAA transistors made of GaAs and InGaAs core. We compute the phonon-limited mobility and the impact of the scattering on the drain current.

The whole paper is divided into four sections. First we explore the effect of scattering in Si GAA nanowire transistors. The phonon-limited mobility and the current reduction due to electron phonon scattering are calculated for a variety of cross-sections. The impact of the real part of phonon scattering self-energies on the current and on the mobility in a narrow Si nanowire transistor has been quantified in section two. A complementary "Appendix" to section two discusses more technically some theoretical issues associated with causality, locality and the real part of the self-energy. In section three, the local power dissipation in a nanowire transistor is computed. The effect of the source region enlargement on the distribution of the carrier energy is also studied. Two different cross-sections are considered ($2.2 \times 2.2\,\mathrm{nm}^2$, $3.6 \times 3.6\,\mathrm{nm}^2$). Different combinations of source, channel and drain length are explored.

Finally, in section four, the effect of scattering in GaAs and InGaAs (GAA) nanowire transistors is investigated. For two cross-sections ($2.2 \times 2.2\,\mathrm{nm}^2$ and $4.2 \times 4.2\,\mathrm{nm}^2$), the effect of

confinement, reduction in the current due to electron–phonon scattering and phonon-limited mobility are discussed.

2 Electron–phonon scattering in GAA Si nanowires (low and high drain bias)

In this work, the electron is described using an effective mass approximation [6,26,27] and a mean field approximation. However, the masses are extracted from tight-binding calculations. Our model used an anisotropic mass tensor [4] and coupled mode space approach, in order to reduce computational time. The phonons are assumed to be in equilibrium, and the electron phonon scattering parameters are from reference [7], which has been shown to be a good approximation for confined structures. Acoustic and f-g-intervalley and intravalley phonon scattering mechanisms have been included. Unless otherwise specified, a gate-all-around (GAA) nanowire field effect transistor structure is considered. The channel is usually assumed to be undoped, and the n-type source and drain region are doped by $10^{20}\,\mathrm{cm}^{-3}$.

In the first part of this section we present the calculations for phonon-limited mobility for different cross-sections (2.2×2.2, 3.2×3.2, 4.2×4.2, 5.2×5.2 and 6.2×6.2 nm^2) [4]. The lengths of the source and drain are 15 nm. In order to extract the mobility we have calculated the current voltage characteristics for different channel lengths 20, 30 and 40 nm. The source and drain resistance is calculated using a regression technique [4]. The channel resistance R_{ch} is computed by subtracting the source and drain resistance $R_{\mathrm{s/d}}$ from the total resistance R_{total}, i.e., $R_{\mathrm{ch}} = R_{\mathrm{total}} - R_{\mathrm{s/d}}$. The phonon-limited mobility is given by:

$$\mu_{\mathrm{pho}} = \frac{L_{\mathrm{ch}}}{R_{\mathrm{ch}} Q_{\mathrm{ch}}} \tag{1}$$

where L_{ch} is the channel length and Q_{ch} is the channel charge by unit length. The calculated mobilities as a function of Q_{ch} for all the cross-sections considered are shown in Fig. 1.

The mobilities decreased when the cross-section decreased. This fact is well known [5]. The reason lies in the dependence of the self-energies on the form factors. The general expression for the form factor is given in [10]. However, in the uncoupled mode approximation it reduces to the following expression:

$$\Sigma_{\nu}^{<n,n}(x) = \sum_{m,\nu 1} C_{m,\nu 1}^{n,\nu} G_{\nu 1}^{<m,m}(x)\, F_{m,\nu 1}^{n,\nu}(x) \tag{2}$$

where the $C_{m,\nu 1}^{n,\nu}$ are the electron–phonon coupling between $(\nu 1, m)$ and (ν, n) states, n, m represent modes, and $\nu 1$, ν represent valleys. $G_{\nu 1}^{<m,m}(x)$ is the projection of the lesser Green function to the direction of transport x-axis. The $F_{m,\nu 1}^{n,\nu}(x)$ are the form factors given by:

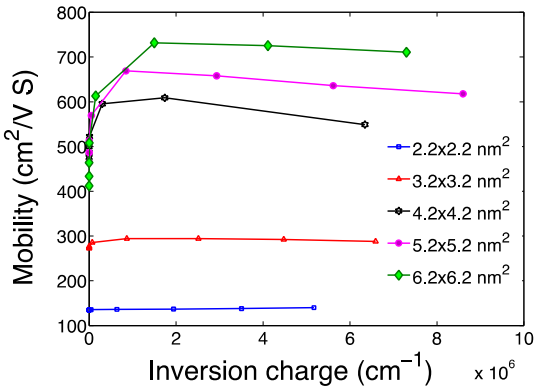

Fig. 1 Phonon-limited mobility as a function of the inversion charge for different cross-sectional nanowires

$$F_{m,\nu 1}^{n,\nu}(x) = \int \mathrm{d}y\mathrm{d}z \left|\Psi_{x,\nu}^{n}(y,z)\right|^2 \left|\Psi_{x,\nu 1}^{m}(y,z)\right|^2 \tag{3}$$

where $\left|\Psi_{x,\nu}^{n}(y,z)\right|^2$ is the wave function squared for the cross-section at the x-location. The integration is over the whole cross-section. These form factors are proportional to the inverse of the cross-sectional area. Note that wave functions are normalized in the cross-sectional area. Figure 1 shows a slight decrease in mobility as the inversion charge increases, and this results from increasing intersub-band scattering at high fields [4]. It is worth pointing out that our results are in agreement with previous works [7,28–31] using different physical models. The agreement of our results with full-band method (tight-binding) is a consequence of the similarity of the density of states between the confined phonons and bulk phonons. This has been pointed out in a previous paper [4].

Another important aspect is the impact of phonon scattering in the electron current at high drain bias. This impact is measured by the ballisticity coefficient defined by $100\times(I_{\mathrm{scatt}}/I_{\mathrm{bal}})$ and by the percentage of the current reduction due to scattering, $100*(I_{\mathrm{bal}}-I_{\mathrm{scatt}})/I_{\mathrm{bal}}$. The current reduction for the $2.2\times2.2\,\mathrm{nm}^2$ cross-section as a function of the channel length is shown in Fig. 2. The current reduction considering only scattering in the channel is also shown for comparison. As expected as the channel increases, the effect of scattering increases becoming 85 % for 40 nm channel length. An important finding is that the impact of scattering for the 6 nm channel length small device is larger than 50 %. For the device of $6.2\times6.2\,\mathrm{nm}^2$ at 6 nm channel length the current reduction is still substantial and is equal to 32 %, and the corresponding value for a 40 nm channel is 50 %.

We have demonstrated that phonon scattering is certainly non-negligible for *small* nanowire transistors. In addition the

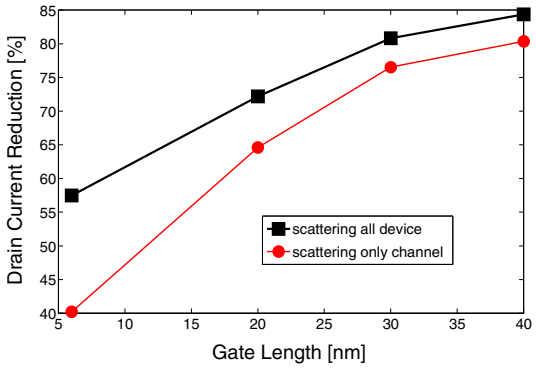

Fig. 2 Current reduction for the 2.2×2.2 nm^2 cross-section at different channel lengths. The device with scattering only in the channel is also showed for comparison

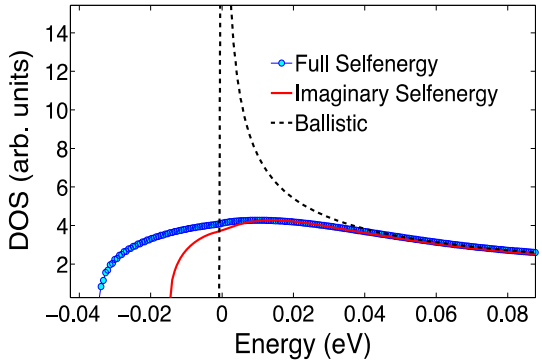

Fig. 3 The density of states for the ballistic case, the case where the full self-energy is included and the case when only the imaginary part is considered

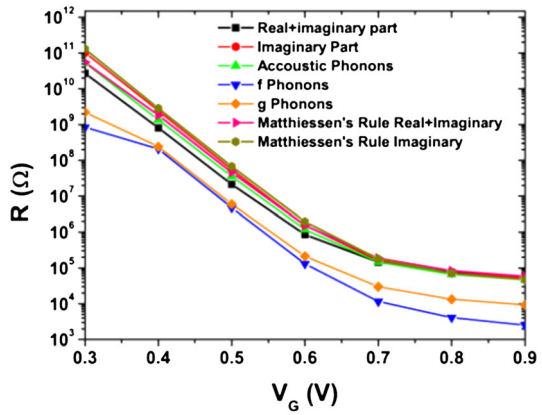

Fig. 4 Resistance of a GAA nanowire transistor of $2.2 \times 2.2\,\text{nm}^2$ cross-section. The resistances associated with each individual phonon mechanism and with/without real part of phonon self-energy are shown

reduction in the nanowire cross-section makes the impact of scattering more severe.

3 Impact of the real part of the phonon scattering self-energies on silicon nanowires

As mentioned in the introduction, only a few authors have studied [10,11] the impact of the real part of phonon self-energy and have concluded that neglecting it produces an underestimate of electron density.

A violation of causality or more general sum rules [32] can be associated with a breakdown of charge conservation [33] or other conservations laws. This is shown in the work of Friedel [34].

From a device point of view the lack of conservation of density states (DOS) and the underestimation of charge density can prevent the accurate determination of electron current. However, the real part of the self-energy also produces a negative local shift in the electron energy that when combined with the enforcement of charge neutrality (Poisson's equation) produces effects in the current and mobility which are difficult to predict without the help of calculations. Technical details of causality and locality issues with self-energies are described in "Appendix."

It is possible with a 1D model and a single Einstein phonon mode [21] to show that there is a clear underestimation of the DOS. This is described in Appendix Section "A simple Einstein model for the self energy" for the case of a strong electron–phonon interaction to illustrate the effects. Figure 3 shows a more general computation for the density of states for a ballistic case and the cases where the real part of the self-energy is included and is neglected for the simple model with more realistic parameters. The integrated DOS for the case with only the imaginary part is 10 % smaller than the full self-energy case. The two most relevant features are (i) a decrease of the DOS when the real part is neglected and (ii) a downward shift in energy when the real part is considered. The ballistic DOS shows the clear 1D character.

Studying the impact of the full self-energy in real nanostructures is more complex, and here we address some of the issues by highlighting the main findings of our previous works [12,13,35]. In nanowire transistors the neglect of the real part produces an overestimate of the resistance as can be seen in Fig. 4 [13]. The overestimation of the resistance using only the imaginary part is also shown. In addition the contribution of each phonon mechanism is also presented. Acoustic phonon mechanism seems to be the largest contributor to the resistance.

In order to investigate the effect of the real part of the self-energy in GAA Si nanowire transistors we have calculated the transfer characteristic for a device with a 2.2×2.2 nm^2 cross-section. The source, gate and drain are 15, 20 and 15 nm, respectively. The drain bias is assumed 1 mV as we intend to calculate the mobility. The doping in the source and drain is 10^{20} cm^{-3}, and in the channel, the acceptor concentration is equal to 10^{16} cm^{-3}.

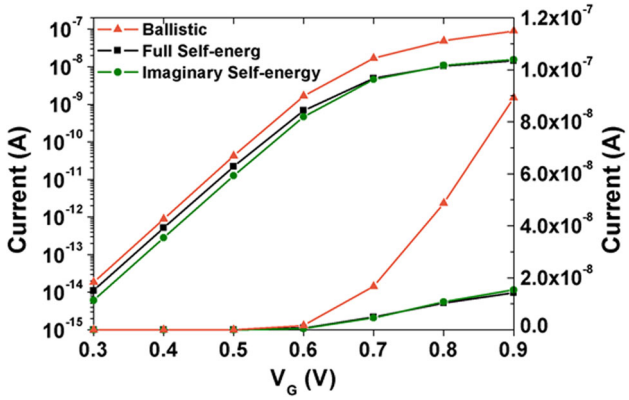

Fig. 5 Current–voltage characteristic for a device with only the imaginary self-energy (*circles*) and with the full self-energy (*square box*)

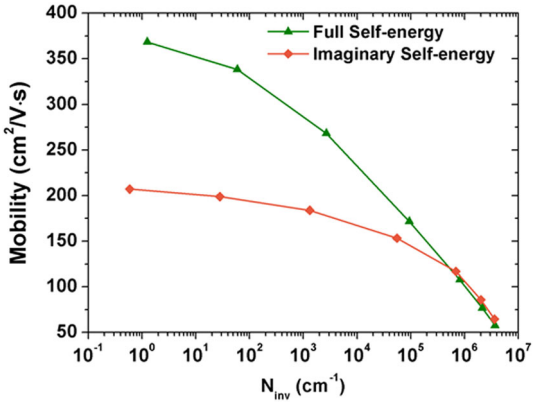

Fig. 7 Phonon-limited mobility calculated for the two cases of Fig. 5

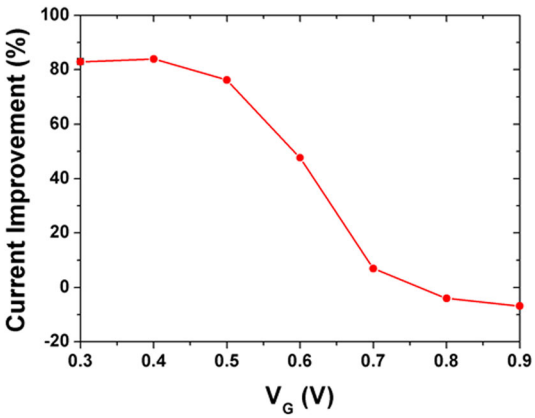

Fig. 6 Percentage increase in the current when the real part of the self-energy is added

Figure 5 compares the drain current versus gate bias for the cases with the full self-energy and the case where the imaginary part is considered. At low gate bias the case with the full self-energy produces a large current but as the gate bias increases the difference between the currents of the two cases becomes smaller. The reason is a subtle one. As we mentioned before the electron density and density of states are larger in the case of the full self-energy, and therefore we will expect a large current in the case of the full self-energy. This is true when the effect on the source–drain barrier of the inversion charge in the channel is negligible. This happens at low gate bias. However, as the charge of the channel increases, the height of the barrier decreases slowly in the full self-energy case. This effect tends to decrease the difference between the two currents while increasing gate bias. This is more evident in Fig. 6, which shows the current increase percentage, when the real part is incorporated. A more quantitative explanation using the density of states and the current spectra can be found in Ref [13].

Now we turn to the impact of the real part of the self-energy in the calculation of the mobility. Figure 7 compares the mobility vs inversion charge for the cases with the full self-energy and the case where the imaginary part is considered. The mobility calculation in this section used a slightly different approach than the first section; here the resistance of the transistor is calculated by subtracting the ballistic resistance (Schur's model [13]) at each gate bias. This approach is more suitable for transistors at all gate bias. This extracted mobility better reflects the effect of scattering in the device and follows the trend of current reduction due to scattering, as the ballistic resistance of the whole transistor is subtracted. From the figure, it can be seen that the mobility reduces as the charge increases with increasing gate bias. This is because the impact of scattering increases with increasing gate bias. The mobility is larger in the case of the full self-energy. The reason is quite similar to the current comparison. Using Eq. 1 of Sect. 1, we can see that a large mobility, in the case of the full self-energy, is a result of a strong drop in resistance even if the inversion charge is large. This decrease in resistance (or increase in the current $I = V/R$) undoes the effect of the inversion charge. The carrier velocity in the case of the real self-energy is large, due to a downward shift in the electron energy (caused by the real part of the self-energy). This shift acts like a negative potential increasing the electron k-vector and its contribution to the current. However, when increasing gate bias, the large charge in the channel in the full self-energy case produces (through an electrostatic self-consistent loop) an upward shift in the source–drain barrier. This upward shift tries to undo the downward shift making the charges of both cases rather similar. This reduces the difference between the two mobilities. In the above effect, the source and drain are partially implicated. We should note that this effective shift downward in the electron self-energy appears as an increase of tunneling under the energy sub-band [13].

We have also calculated the current–voltage characteristic for a long (50 nm) Si nanowire surrounded by a gate and

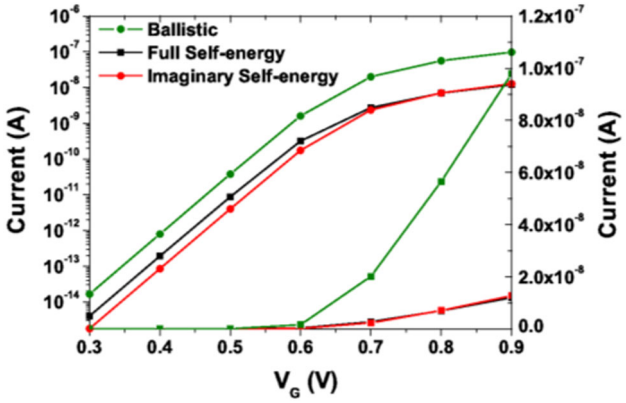

Fig. 8 Current characteristic of a nanowire (not a MOSFET) comparing a case with full and imaginary self-energy

with a similar cross-section to that of the previous nanowire transistor. The currents in the cases with and without real self-energy are shown in Fig. 8. When comparing both currents, the same trend as in the Si nanowire transistor previously considered is shown.

In summary, we have studied the impact of the real part of the self-energy on the current–voltage characteristic and on the mobility of a narrow Si GAA nanowire transistor. It has also been shown that the same behavior follows for a gated nanowire, so the effect is not strongly related to the source and drain regions. In general, we have confirmed that the neglect of the real part of the self-energy underestimated the electron density in the channel but this effect tends to disappear as the channel charge increases. This is due to the increasing impact of the channel charge in the channel electrostatic, which modifies the source–drain barrier.

4 Power dissipation in silicon nanowire transistors

As mentioned in the introduction, electrons in small transistors experience large applied bias within a distance of a few nanometers. A drain bias, as large as 1 V, is expected to be applied in these devices. This large change in electron potential energy leaves strong non-equilibrium electron distribution inside devices. In small cross-sections confinement and a reduced density of states make the selection of final energy states more difficult and also the relaxation of hot electrons. The fact that optical phonons in Si have an average energy of less than 100 meV implies that several phonon processes are needed to cool down a 0.4 eV hot electron. All the above effects decrease energy relaxation; however, in small cross-sections, the effective electron–phonon coupling is largely due to the form factor increasing [5], and this consequently increases energy relaxation. This interplay between different factors makes electron relaxation in nanotransistors a complex phenomena. Usually the Joule power relation $I * V(r)$ [37] is used to estimate local energy dissipation; however, for very small transistors a more detailed analysis is required.

In this section a study of local power dissipation in a Si GAA nanowire transistor will be carried out. We use the same NEGF methodology, effective mass Hamiltonian and phonon scattering mechanism as in the previous sections.

As mentioned before we have assumed that the phonons are in equilibrium (300 K). It is known that the inclusion of non-equilibrium phonons results in a decrease of the current. However, the modeling of non-equilibrium phonons is a difficult task for small nanowires. Recent calculations of self-heating using atomistic phonons show a decrease in the current [36].

In order to make this account more complete the relevant equations needed to understand the energy exchange between the electron and phonon systems in the NEGF formalism are presented. The electron energy current can be expressed as:

$$J_{\text{Energy}}(r) = \int \varepsilon J(r, e) \mathrm{d}e \tag{4}$$

where $J(r, e)$ is the local electron current for energy e. The local power transfer by the electron to the lattice is given by:

$$P(r) = -\nabla \cdot J_{\text{Energy}}(r) \tag{5}$$

Equation 4 can be written in a more transparent way explicitly showing the Joule power component [22,37]:

$$P(r) = -\nabla \cdot J_K(r) + E(r) \cdot J(r) \tag{6}$$

This equation represents the total local power dissipated by the electron system into the phonon system. The last term in the RHS is the local Joule power, which when integrated through the device produces the well-known Joule power $I * V$. The first term in the RHS of (6) is the change in the "kinetic energy" current of the electron ensemble, if this kinetic energy current does not change it means that the electron system is following the bending of the potential and therefore not increasing the electron mean energy relative to the conduction band minima, i.e., no hot electron phenomena. In this case the electron system or more specifically the electrons taking part in the current are in equilibrium with the local potential. On the other hand a large change in kinetic current implies the heating or cooling of the electron system. Using Eqs. 11b and 12b in the appendix, we can obtain the detailed balance equation [22,23]:

$$\begin{aligned} -\nabla \cdot J_{\text{Energy}}(r) &= P(r) \\ &= \int e \left\{ G^{<}(r, r, e) \Sigma^{>}(r, r, e) \right. \\ &\quad \left. - G^{>}(r, r, e) \Sigma^{<}(r, r, e) \right\} \mathrm{d}e \end{aligned} \tag{7}$$

In deriving Eq. 7, local electron–phonon self-energies have been assumed. The electron phonon model used in this paper is also local. Equation 7 shows that the energy mismatch between the rate of ε-outgoing (leaving the state of energy ε) electrons and the rate of ε-outgoing holes is given to the phonon system. Note that for elastic scattering the expression in the curly brackets becomes zero [21,22]. Equation 6 in this paper is used to estimate the power dissipated in the electrons inside the device. In reference [21], the same formalism was applied to study power dissipation in a resonant tunneling structure. The calculations presented here have relevance for 3D nanotransistors with a sub-10-nm channel length, but they can easily be extended to longer channel lengths as the electrons leaving the drain will be hot electrons in a 20 nm drain region.

We have simulated a variety of source, channel and drain lengths in order to see the role of different device regions in power dissipation or energy transfer between the electron and phonon systems. Two different cross-sections have been investigated: a $2.2 \times 2.2\,\mathrm{nm}^2$ and a $3.6 \times 3.6\,\mathrm{nm}^2$.

First we concentrated on high gate bias for which power dissipation is relevant, and finally, we will present some effects at low gate bias, such as the cooling of the source, which is related to the Peltier effects [22].

Figure 9a, b shows the current spectra and local power dissipation along the nanowire axis for a $2.2 \times 2.2\,\mathrm{nm}^2$ with a 10/6/24 nm for source/gate/drain. In Fig. 9b, the divergence of the kinetic energy and the local Joule power are shown as a comparison.

The average current energy [22] and the first sub-band are also shown in Fig. 9a. The average current energy decreases over distance within the drain as expected, due to the energy relaxation of hot electrons as they move through the drain. However, note that even after 24 nm the electron current is not yet in equilibrium with the lattice and at the end of the source the current mean energy has dropped by half (200 meV) of the 400-meV potential drop. Therefore, the power dissipated inside the device cannot be solely described by the Joule term in Eq. 6. This is quantitatively demonstrated in Fig. 9a, which shows the density of the total power, the Joule power and divergence of the kinetic energy current along the nanowire transistor at $V_G = 0.9\,\mathrm{V}$. The Joule power essentially follows the derivative of the conduction band energy, and its maximum is at $x = 16.4\,\mathrm{nm}$ with a value of $2.79 \times 10^{-7}\,\mathrm{W/nm}$. However, the maximum of the total power dissipated is a little bit farther ($x = 18\,\mathrm{nm}$) and its value is $2.74 \times 10^{-8}\,\mathrm{W/nm}$, ten times smaller than the Joule value. As previously pointed out, the kinetic and Joule terms are quite similar and their subtraction produces the total power. This subtraction is quite sensitive and requires a high degree of convergence in the solution of the nonlinear NEGF equation in the presence of scattering.

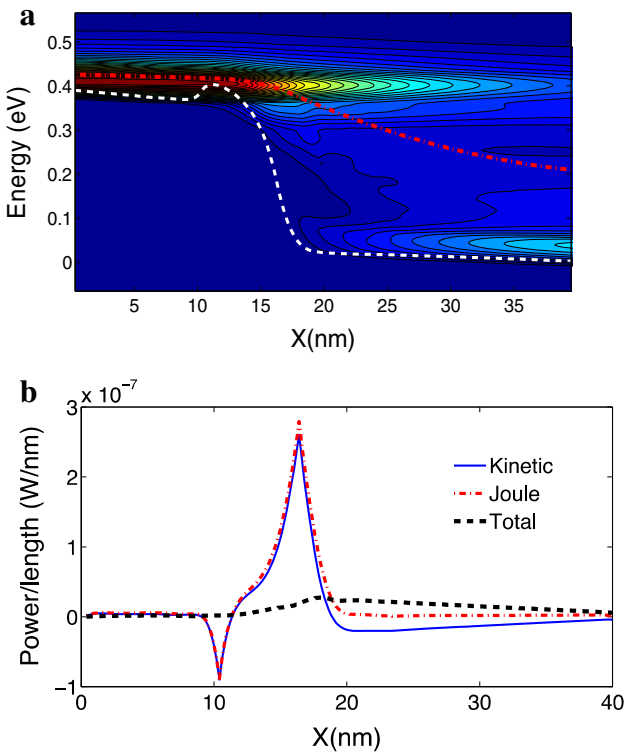

Fig. 9 **a** Current spectra for the silicon nanowire transistor with $2.2 \times 2.2\,\mathrm{nm}^2$ cross-section at $V_G = 0.9\,\mathrm{V}$. The first sub-band (*white dotted line*) and the average current energy (*red dotted line*) are also presented. **b** Spatial distribution of the power per unit length. The Joule power (*red dashed*) and the divergence of the kinetic energy (*blue*) are also shown

Figure 10a, b shows the current spectra and power per unit length in the device with the larger cross-section. In this case dissipation is not as strong as in the case of the smaller cross-section, as it was expected. The mean energy of the electron current only drops 50 meV at the end of the drain, making the electron system very far from equilibrium inside the device. The position shifting between the maxima of the total and Joule power density is equal to (21.4–16.4) nm = 5 nm. The maximum values of the total and Joule power density by unit length are 1.22×10^{-6} and $4.04 \times 10^{-8}\,\mathrm{W/nm}$. It is worth noting that the ratio between the two values is two orders of magnitude. As mentioned before, the reason for this large ratio when compared to the corresponding value of the small cross-sectional device is the large phonon scattering rate in the small cross-sectional device. The corresponding integrated total and Joule power through the whole device is 6.3538×10^{-7} and $3.2667 \times 10^{-6}\,\mathrm{W/nm}$, respectively. In order to further explore relaxation in the drain region, devices with 10-/6-/64 nm source/channel/drain regions were simulated.

Figure 11 shows the current spectra for the small and large cross-sections, respectively. The average current energy flattens at the drain end, showing that equilibrium has been

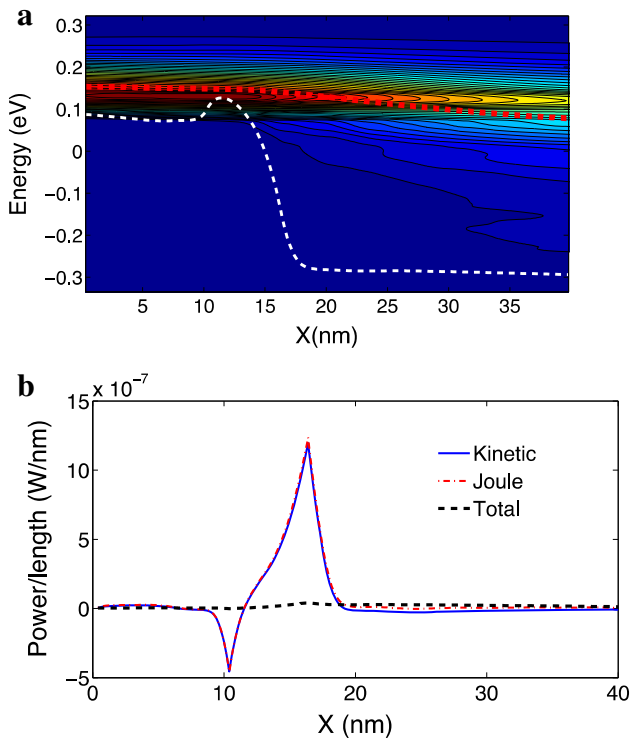

Fig. 10 **a** Current spectra for the silicon nanowire transistor with $3.6 \times 3.6\,\text{nm}^2$ cross-section at $V_G = 0.9\,\text{V}$. The first sub-band (*white dotted line*) and the average current energy (*red dotted line*) are also presented. **b** Spatial distribution of the power per unit length for the case of Fig. 3. The Joule power (*red dashed*) and the divergence of the kinetic energy (*blue*) are also shown

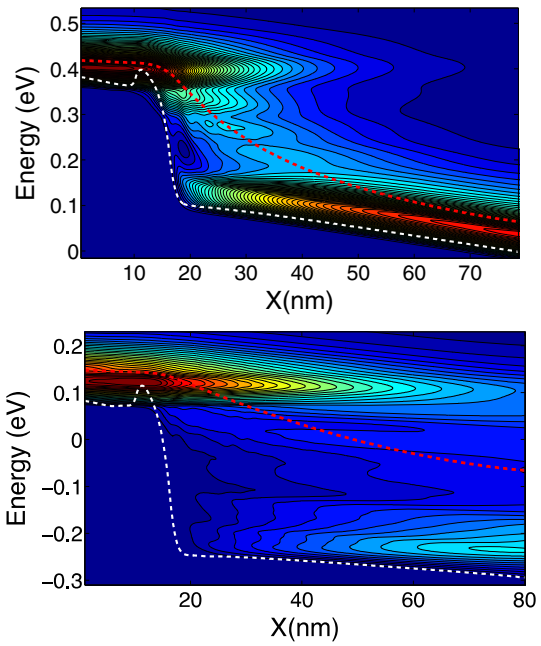

Fig. 11 Current spectra for the transistors with $2.2 \times 2.2\,\text{nm}^2$ (*upper panel*) and 3.6×3.6 nm^2 (*lower panel*) cross-sections and 64-nm drain length at $V_G = 0.9\,\text{V}$ The first sub-band (*white dotted line*) and the average current energy (*red dotted line*) are also presented

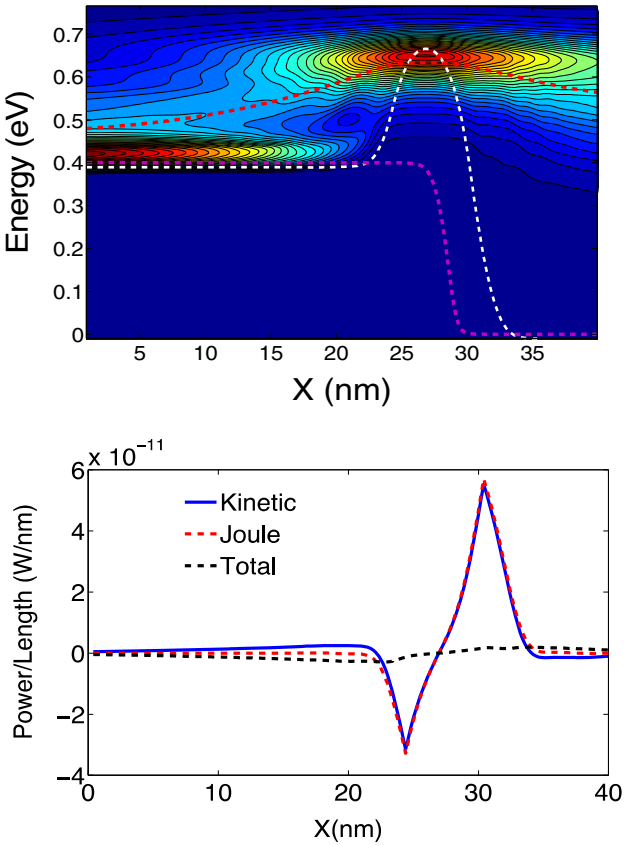

Fig. 12 *Upper panel* current spectra for the transistor with $2.2 \times 2.2\,\text{nm}^2$ cross-section and 24-nm source length at $V_G = 0.4\,\text{V}$. The first sub-band (*white dotted line*) and the average current energy (*red dotted line*) are also presented. The quasi-Fermi level is shown in magenta. *Note* that the difference between the average current energy and the Fermi level is equal to the Peltier coefficient. *Lower panel* spatial distribution of the power per unit length for the case of Fig. 3. The Joule power (*red dashed*) and the divergence of the kinetic energy (*blue*) are also shown

reached. However, this is not the case for the large cross-sectional device. The average current energy for the large cross-sectional device is still changing at the contact. The total power dissipated inside the device is 56 % of the Joule power. It is important to highlight that even for a relatively long drain region (65 nm), the hot electrons in the $3.6 \times 3.6\,\text{nm}^2$ device leave the drain without relaxation.

The value of the total and Joule reductions integrated through the whole simulated device length is summarized in Table 1 at $V_G = 0.9\,\text{V}$. In what follows, we highlight and discuss the relevant issues for every individual configuration.

In the small cross-sectional device, in which source/channel/drain lengths are equal to 10/6/64 nm, hot electrons completely relax in the drain. At $V_G = 0.9\,\text{V}$, the power dissipated is 92 % of the Joule power.

Longer source devices with 24-/6-/10 nm long source/channel/drain regions and $2.2 \times 2.2\,\text{nm}^2$ and $3.6 \times 3.6\,\text{nm}^2$ cross-section have been studied. Figure 12 shows the current

Table 1 Current reduction and percentage of joule power dissipated in different dimension devices

Source/channel/ drain (nm)	Cross-section (nm^2)	P_T/P_J (%)	Current reduction (%)
10/6/24	2.2×2.2	50	50
10/6/24	3.6×3.6	19	30
24/6/10	2.2×2.2	26	57
24/6/10	3.6×3.6	11	22
10/6/64	2.2×2.2	92	63
10/6/64	3.6×3.6	56	38
10/40/30	2.2×2.2	73	82
10/40/30	3.6×3.6	36	63

spectra and the local power at low gate bias ($V_G = 0.4$ V) in the small cross-sectional device. In this case the gate barrier energy is around 250 meV; source electrons surmounting the barrier through phonon absorption cool the source of the transistor. The upper panel of Fig. 12 shows that the average current energy in the drain is larger than the drain indicating a net cooling of the lattice. The integrated power inside the whole simulated device region extracted from Fig. 8 is negative, as the positive power dissipated in the drain is unable to compensate the cooling of the source. The Joule power and integrated power are 1.0349×10^{-10} and -2.2232×10^{-11} W, but the total power integrated, including the device and its surrounding equilibrium regions, is positive and equals the Joule power as expected. This is similar to what happens at the interface of regions with different Peltier coefficients [22]. This negative local power occurs at very low gate bias. At high gate bias, the source cooling is very small as compared to the power dissipated in the drain, even for short drain extensions. At $V_G = 0.9$ V, the joule and total powers are 6.8285×10^{-7} W and 1.8411×10^{-7} W, which results in only 26 % of the joule power dissipated inside the device. As the length of these devices is similar to the 10-/6-/24-nm device a comparison between the relaxation in the source and drain can be established.

Table 1 presents the current reduction and the percentage of Joule power (i.e., $100 \times P_T/P_J$, where P_T is the power dissipated inside the device given by the integration over the whole device of the RHS of Eq. 6 or 7 and P_J is the integration of the second term of Eq. 6) dissipated inside the device for different combinations of source/channel/drain dimensions.

Two features stand out: (i) Current reduction is mainly related to the channel length and in the small/large cross-section is approximately equal to 55/30 % on average for 6-nm channel length, (ii) total power dissipation is mostly related to the drain size, but it also increases with channel length. Devices with 10/40/30 and 10/6/64 have the same channel+drain length; however, the 64 nm drain device has a 20/37 % large power dissipation than the 30 nm drain in the small/large cross-section, respectively.

In this section local power dissipation through a narrow GAA nanowire transistor has been extracted for two different cross-sections. The electron system in the smaller cross-sectional transistor requires more than 60 nm to release its energy at the drain. For the large cross-sectional device more than 80 nm would be needed. This is an important finding as it means that most of the transistor's power will be dissipated outside its active region and deep into the drain–contact interface. This is true for relatively large cross-sections but as the transistor's cross-section shrinks, the power dissipated close to the channel–drain interface increases. It is worth noting that interface roughness and impurity scattering are elastic mechanisms and do not contribute to power dissipation. We have used bulk phonons, but our previous work suggests than even if this is a crude approximation for very small cross-sections, the phonon density of states for confined phonons is still concentrated around the bulk phonon energies see ref. [4]. Finally, we have observed as pointed out by other authors [21,22] that a cooling of the lattice occurs in the source region at low gate bias.

5 GaAs and InGaAs nanowire field effect transistors

In this section, III–V (GaAs [100] and InGaAs [100]) n-channel nanowire field effect transistors (NWFETs) are discussed. III–V nanowires are under consideration for use in NWFETs because of their high mobility compared to Si. NWFETs of cross-sections 2.2×2.2 nm^2 and 4.2×4.2 nm^2 and channel length 6 nm have been simulated. The source and drain are 15 nm long and are doped at 10^{20} cm^{-3}. The channel is undoped. The structure of the NWFETs is given in Fig. 13. We have calculated the transfer characteristic for a drain bias of 0.6 V and extracted the current reduction. In addition, we have extracted the mobility using the methodology outlined in Sect. 3. The Hamiltonian has been written in the effective mass approximation; the masses are extracted from tight-binding calculations [38]. These masses calculated from tight-binding have been extracted as a function of the cross-section. The valleys are assumed to be isotropic.

Scattering of elastic acoustic phonons, optical intervalley ($\Gamma - L$, $L - X$, $\Gamma - X$, $L - L$, $X - X$), intravalley and polar

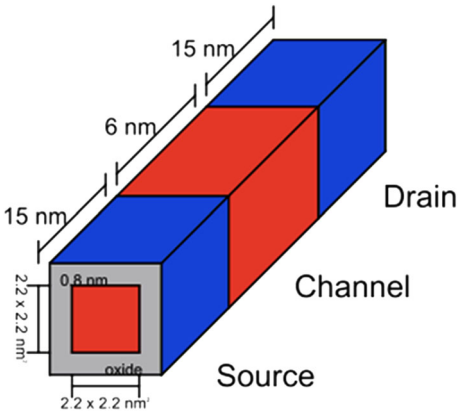

Fig. 13 Example of the structure of a $2.2 \times 2.2\,\text{nm}^2$ cross-sectional, 6-nm channel NWFET

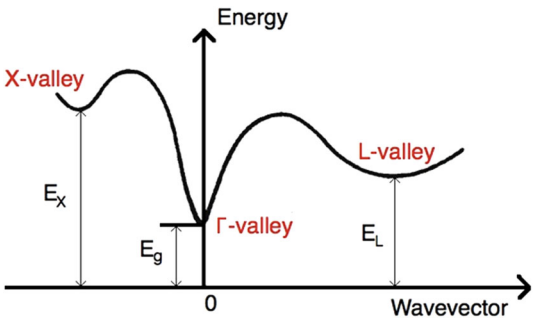

Fig. 14 Schematic of the band structure of GaAs and InGaAs in bulk

optical (intravalley) phonons (POP) have been considered. In GaAs and InGaAs, polar optical phonon scattering is the dominant scattering mechanism, followed by $\Gamma - L$ intervalley scattering. The polar phonon scattering self-energy has been considered to be local [39]. This approximation is made here as the implementation of non-local self-energies will render the use of the recursive algorithm invalid. The scattering parameters have been taken from [40].

Figure 14 shows a schematic of the band structure for bulk GaAs and InGaAs. In order of increasing energy the valleys are Γ, L and X. For the cross-sections of the devices simulated in this paper, there is strong confinement, which in the $2.2 \times 2.2\,\text{nm}^2$ cross-sectional devices causes the low mass Γ-valley to become elevated in energy [41]. This produces low current and mobility in the $2.2 \times 2.2\ \text{nm}^2$ devices.

Figures 15 and 16 show the energy-resolved current spectra for the GaAs and InGaAs core, $2.2 \times 2.2\,\text{nm}^2$ cross-sectional NWFETs, respectively, with the position of the valleys. In both the GaAs and InGaAs core device the Γ-valley has become elevated in energy. The Γ-valley is lower in InGaAs in spite of the lower effective masses (in comparison with GaAs) because of the greater separation between the Γ-valley and the L-valley in bulk InGaAs than GaAs. For the both the GaAs and InGaAs, $4.2 \times 4.2\,\text{nm}^2$ cross-sectional

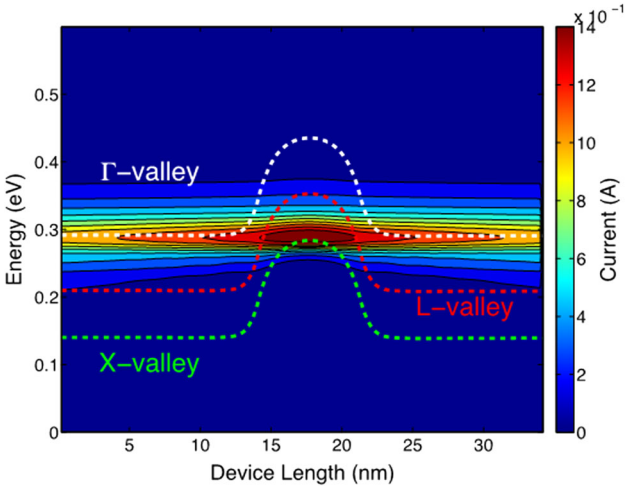

Fig. 15 Energy-resolved current spectra for a $2.2 \times 2.2\,\text{nm}^2$ cross-sectional, 6-nm channel length, GaAs core NWFET

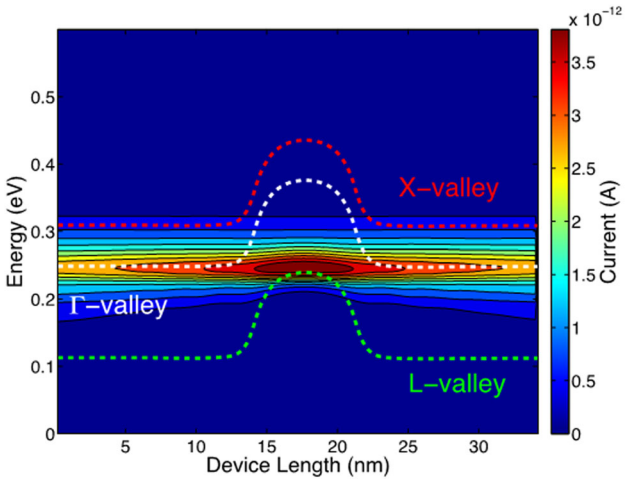

Fig. 16 Energy-resolved current spectra for a $2.2 \times 2.2\,\text{nm}^2$ cross-sectional, 6-nm channel length, InGaAs core NWFET

devices the order of the valleys is the same as bulk: Γ, L then X.

The $I_D - V_G$ characteristics at high drain bias for the $2.2 \times 2.2\,\text{nm}^2$ cross-sectional, 6 nm channel GaAs and InGaAs devices are shown in Fig. 17. There is higher current in the GaAs NWFET because the density of states is higher. This is caused by both the high mass L and X-valleys contributing to the current. Figure 18 shows the $I_D - V_G$ characteristics for the $4.2 \times 4.2\,\text{nm}^2$ cross-sectional, 6-nm channel devices at high drain bias. Due to the lower effective masses for the larger cross-sectional device, the current for the GaAs and InGaAs core devices is now closer.

The $I_D - V_G$ characteristics for each valley in the $2.2 \times 2.2\,\text{nm}^2$ cross-sectional, GaAs core NWFET are shown in Fig. 19. The high mass L- and the X-valley provide the largest contribution to the transport; there is little current in

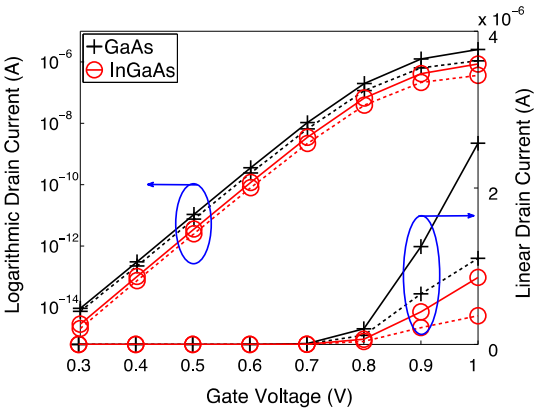

Fig. 17 $I_D - V_G$ characteristics for the $2.2 \times 2.2\,nm^2$ cross-section GaAs and InGaAs core NWFETs at high drain bias $V_D = 0.6\,V$. The *bold* and *dotted lines* represent ballistic and scattering simulations, respectively

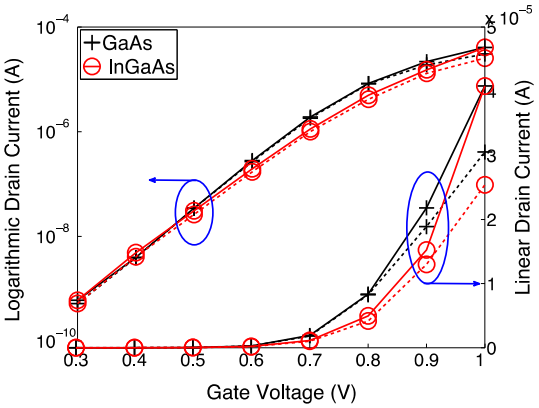

Fig. 18 $I_D - V_G$ characteristics for the $4.2 \times 4.2\,nm^2$ cross-section GaAs and InGaAs core NWFETs at high drain bias $V_D = 0.6\,V$. The *bold* and *dotted lines* represent ballistic and scattering simulations, respectively

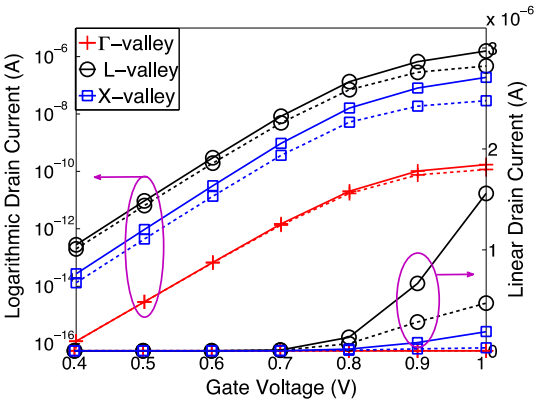

Fig. 19 $I_D - V_G$ characteristics for each valley for the $2.2 \times 2.2\,nm^2$ cross-sectional GaAs core NWFET at high drain bias $V_D = 0.6\,V$. The *bold* and *dotted lines* represent ballistic and scattering simulations, respectively

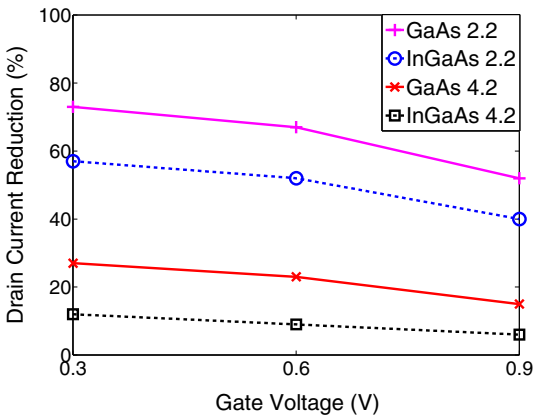

Fig. 20 Percentage reduction in the current due to scattering at high drain bias, $V_D = 0.6\,V$ for all simulated III–V NWFETs

the low mass Γ-valley. For the equivalent InGaAs NWFET, the majority of the current is in the L-valley, with little current in the Γ and X-valleys.

Figure 20 shows the percentage reduction in the current due to scattering for all the simulated III–V NWFETs. It can be observed that the reduction in the current is much smaller for the $4.2 \times 4.2\,nm^2$ cross-sectional devices. This is because the strength of the electron–phonon coupling decreases with increasing cross-section. It is also apparent that the percentage reduction in the current is less for the InGaAs core NWFETs than the corresponding GaAs devices. This is because the electron–phonon coupling is weaker in InGaAs. This is because the deformation potential is higher for the $\Gamma - L$ intervalley scattering in InGaAs, and this is the second most dominant scattering mechanism after POP scattering. POP energy is very similar for GaAs and InGaAs.

The phonon-limited mobility for all the simulated III-V NWFETs is shown in Fig. 21. For the $2.2 \times 2.2\,nm^2$ cross-sectional NWFETs the mobility is very low. This is because of the low mass Γ-valley becoming elevated in energy, which results in low current and mobility. For the $4.2 \times 4.2\ nm^2$ cross-sectional NWFETs, the InGaAs core NWFET has the highest mobility, with a mobility of $1887\,cm^2/Vs$ at low channel charge. For all the NWFETs, the mobility decreases as the channel charge increases; this is due to the scattering increasing as the gate bias increases.

6 Conclusion

We have demonstrated that the effective mass approximation, in combination with the NEGF formalism, is a powerful tool in exploring the trend of different phenomena in small transistors. Quantum effects are described with the simplest approximation, and therefore, the technique is very efficient computationally. However, the masses and other parameters

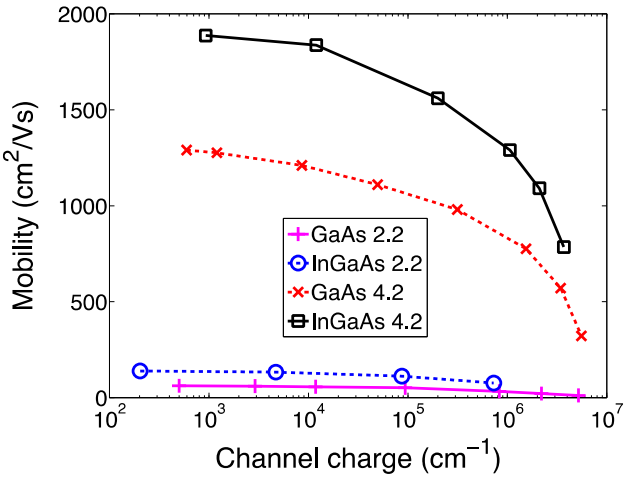

Fig. 21 Phonon-limited mobility for the GaAs and InGaAs, 2.2 × 2.2 nm^2 and 4.2 × 4.2 nm^2 cross-sectional NWFETs

need to be renormalized (calibrated) using more sophisticated physical models such as density functional theory or tight-binding. In principle, non-parabolicity can be considered by introducing adiabatically energy-dependent masses. The inclusion of phonon scattering introduces a heavy computational burden as it adds a new self-consistent cycle. However, as we used local self-energies for scattering, the recursive algorithm can be still used and lightens the computational burden. For polar phonon scattering, the use of local self-energies is a more questionable approximation but it makes the computation possible. The use of different approximations for non-local phonons has been briefly studied in Ref [42]. Recently, a promising new methodology that conserves the current has been suggested [43,44], but it has not been considered in this work.

The impact of scattering, as a function of the cross-section and channel length, on the current and mobility has been thoroughly investigated for different materials. We found that the main increase of scattering with decreasing cross-sections comes from the proportionality of the scattering rate to the cross-sectional area. The impact of the real part of the self-energy on the current voltage characteristic and mobility has also been computed, yielding a 50 % drain current increase at low gate bias, and is negligible at high gate bias. The mobility follows a similar trend. Power dissipation, spatially resolved, has been studied in a combination of source, channel and drain dimensions and for two different cross-sections. Hot electrons entering the drain need at least a 60 nm drain length to thermalize for a 2.2×2.2 nm^2 cross-sectional device at a 0.4-V drain bias. This is for a short 6 nm gate length device. For large channel lengths, electrons start to thermalize in the channel and consequently require less drain length to reach equilibrium. However, electron energy relaxation becomes slower for large cross-sectional devices. This helps with power dissipation, as most of the power will be delivered at the drain–contact interface, which is not surrounded by an oxide material.

Acknowledgments This work was supported by EPSRC under Career Acceleration Fellowship Grant No. EP/I004084/2: Quantum Simulations of Future Solid-State Transistors of A. Martinez.

Appendix

Self-energies in silicon nanowires: causality and locality issues

In this Appendix we gather together various technical matters relating the NEGF theory that are mentioned in the main part of the paper. The NEGF methodology is most efficient for modeling ballistic inhomogeneous transport in nanostructures where a simple artificial self-energy is incorporated to describe the injection and extraction of carriers at the boundaries. Discrete impurity scattering and boundary scattering may be incorporated non-perturbatively via a suitable self-consistent potential. However, for phonon scattering it is necessary to introduce more appropriate physical self-energies within perturbation theory and this raises the complexity of computations very considerably.

Transport properties and Green functions

For a stationary system, neglecting initial state correlations, the transport properties may be determined from two non-equilibrium Green functions, G^R (retarded Green function) and $G^<$ (lesser Green function):

$$N(\mathbf{r};E) = -\frac{1}{\pi}\mathrm{Im}\, G^R(\mathbf{r},\mathbf{r};E) \equiv A(\mathbf{r};E) \tag{8}$$

local density of states (spectral density)

$$n(\mathbf{r};E) = -\frac{2e}{2\pi}\mathrm{Im}\, G^<(\mathbf{r},\mathbf{r};E) \tag{9}$$

electron charge density (with spin degeneracy)

$$\mathbf{J}(\mathbf{r};E) = -e\frac{\hbar}{4\pi}\mathbf{m}^{-1}.(\nabla-\nabla')G^<(\mathbf{r},\mathbf{r};E)\,|_{r\to r'} \tag{10}$$

electron current density

$G^<$ is determined from two coupled equations: first (Dyson's equation) for G^R

$$[E - H(\mathbf{r}_1) + i\eta]G^R(\mathbf{r}_1, \mathbf{r}_2; E) - \int d\mathbf{r}\Sigma^R(\mathbf{r}_1, \mathbf{r}; E)G^R(\mathbf{r}, \mathbf{r}_2; E) = \delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (11a)$$

or, in operator form:

$$[E + i\eta - H - \Sigma^R]G^R = 1 \quad (11b)$$

where $\eta \rightarrow 0^+$ is a causality factor; second, a quantum kinetic equation for $G^<$

$$[E - H(\mathbf{r}_1) + i\eta]G^<(\mathbf{r}_1, \mathbf{r}_2; E) - \int d\mathbf{r}\Sigma^<(\mathbf{r}_1, \mathbf{r}; E)G^<(\mathbf{r}, \mathbf{r}_2; E) = \int d\mathbf{r}\Sigma^<(\mathbf{r}_1, \mathbf{r}; E)G^A(\mathbf{r}, \mathbf{r}_2; E) \quad (12a)$$

or, in operator form

$$G^< = G^R\Sigma^<G^A \quad (12b)$$

Here, the self-energies $\Sigma^K (K =<, R, A = advanced)$ each comprise terms due to the contact regions (S=source and D=drain) and a term due to the electron–phonon interactions (I):

$$\Sigma^K = \Sigma_S^K + \Sigma_D^K + \Sigma_I^K \quad (13)$$

The self-energies are self-consistent functionals of the various Green functions.

The Hamiltonian H is given by the effective mass model Hamiltonian (parameterized from density functional or tight-binding calculations for the nanostructure band structure):

$$H(\mathbf{r}) = -\frac{\hbar^2}{2}\nabla_i.\mathbf{m}_{ij}^{-1}.\nabla_j + V(\mathbf{r}) \quad (14)$$

where $(1/m)_{ij}$ is the reciprocal effective mass tensor and $V(\mathbf{r})$ is the self-consistent electrostatic potential obtained from Poisson's equation.

Causality and analyticity

The complex-valued nature of the carrier–phonon self-energies derives ultimately from the assumed causal response of the system. This may be clarified by considering the full-double time-dependent non-equilibrium retarded Green function

$$G^R(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2) = \theta(t_1 - t_2)[G^>(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2) - G^<(\mathbf{r}_1, \mathbf{r}_2; t_1, t_2)] \quad (15)$$

or, in center-relative time co-ordinates

$$t = t_1 - t_2; \qquad T = \frac{1}{2}(t_1 + t_2) \quad (16)$$

$$G^R(\mathbf{r}_1, \mathbf{r}_2; t, T) = \theta(t)[G^>(\mathbf{r}_1, \mathbf{r}_2; t, T) - G^<(\mathbf{r}_1, \mathbf{r}_2; t, T)] \quad (17)$$

The nomenclature "causal response" refers to condition (15 or 17) which is strictly an antecedence principle for retarded systems (only the past can influence the future: $G^R(t) = 0$ for $t < 0$).

The Green function $G^R(E)$ in the energy domain E is recovered by Fourier transforming over t and noting that G^R is independent of T in the stationary limit. The Titchmarsh theorem shows that $G^R(E)$ is a causal function: that is, an analytic function in the upper half complex energy plane satisfying a Kramers–Kronig relation or Hilbert transform [14]:

$$Re[G^R(E)] = -\frac{P}{\pi}\int_{-\infty}^{\infty}\frac{Im[G^R(E')]}{E - E'}dE' \quad (18)$$

where P denotes principal value.

Similarly, $\Sigma^R(E)$ is often *assumed* to be a causal function having a causal transform satisfying the dispersion relation or Hilbert transform between the level shift Δ^R and the level broadening Γ^R:

$$Re[\Sigma^R(E)] = -\frac{P}{\pi}\int_{-\infty}^{\infty}\frac{Im[\Sigma^R(E')]}{E - E'}dE' \quad (19)$$

where we define

$$\Delta^R = Re[\Sigma^R] \equiv \frac{1}{2}(\Sigma^R + \Sigma^A) = \Delta^A; \Gamma^R = -Im[\Sigma^R] \equiv -\frac{1}{2i}(\Sigma^R - \Sigma^A) = -\Gamma^A \quad (20a)$$

$$\Sigma^R = \Delta^R - i\Gamma^R \quad (20b)$$

Here A signifies the advanced self-energy (obtained from the retarded form by the conjugate). From the identities,

$$\Sigma^R - \Sigma^A = \Sigma^< - \Sigma^> \quad (21)$$

the scattering function or broadening function may be obtained from

$$\Gamma^R(\mathbf{r}_1, \mathbf{r}_2; E) = \frac{i}{2}[\Sigma^<(\mathbf{r}_1, \mathbf{r}_2; E) - \Sigma^>(\mathbf{r}_1, \mathbf{r}_2; E)] \quad (22)$$

Because G^R is analytic in the upper half complex energy plane it follows that the scattering function Γ^R is positive valued. The dispersion relation (or Kramers–Kronig relation)

leads to the important computational equation

$$\Delta^R(\mathbf{r}_1,\mathbf{r}_2;E)] = \frac{P}{\pi}\int_{-\infty}^{\infty}\frac{\Gamma^R(\mathbf{r}_1,\mathbf{r}_2;E')]}{E-E'}\mathrm{d}E' \tag{23}$$

It is important to note that if there are bound states or resonances in the electron spectrum, then $\Sigma^R(E)$ need not be a simple causal transform (various subtracted forms are required).

Spectral density sum rule

For a non-interacting single-electron system with Hamiltonian$H$, the retarded Green function is simply (in operator form)

$$G_0^R(E) = (E - H_0 + i\eta)^{-1} \tag{24}$$

In a representation that diagonalizes H_0 expression (24) has poles at $E = E_n - i0^+$ *(i.e., in the lower half energy plane)* where E_n are eigenvalues of H_0.

The spectral density operator for the non-interacting system is defined as

$$\begin{aligned} A_0(E) = \delta(E-H_0) &\to -\frac{1}{\pi}Im[G_0^R(E)] \\ &= \frac{-1}{2\pi i}[(E-H_0+i\eta)^{-1} - (E-H_0-i\eta)^{-1}] \end{aligned} \tag{25}$$

For an interacting system the spectral density operator generalizes to:

$$\begin{aligned} A(E) = \frac{-1}{2\pi i}&[(E-H-\Delta^R+i\Gamma^R+i\eta)^{-1} \\ &-(E-H-\Delta^R-i\Gamma^R-i\eta)^{-1}] \end{aligned} \tag{26}$$

A rigorous many-body analysis (Lehmann theorem) shows that the spectral density satisfies a sum rule (see [45] for a simpler discussion):

$$\int A(E)\mathrm{d}E = \mathbf{1} \tag{27}$$

It may be shown [32] that if $\Sigma^R(E)$ or $\Sigma^R(E)G^R(E)$ are causal transforms, then the spectral sum rule (27) is automatically satisfied. Any violations of causality in $\Sigma^R(E)$ or $\Sigma^R(E)G^R(E)$ give rise to errors in the density of states. Examples of the pitfalls are given in [32,35].

Self-consistent Born approximation

The local conservation laws (including current conservation, energy conservation, momentum conservation) are enforced by the Ward identities [45,46]. At lowest order in perturbation theory, the Ward identities require that the self-energies must satisfy the self-consistent Born approximation (SCBA). For example, for a generic inelastic phonon scattering with phonon energy $\hbar\omega_q$ the SCBA has the form (dropping the label I):

$$\begin{aligned} \Sigma^<(\mathbf{r}_1,\mathbf{r}_2;E) = \int &\frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}|C(\mathbf{q})|^2 \mathrm{e}^{i\mathbf{q}.(\mathbf{r}_1-\mathbf{r}_2)} \\ &\{(N_q+1)G^<(\mathbf{r}_1,\mathbf{r}_2;E+\hbar\omega_q) \\ &+N_qG^<(\mathbf{r}_1,\mathbf{r}_2;E-\hbar\omega_q)\} \end{aligned} \tag{28a}$$

$$\begin{aligned} \Sigma^>(\mathbf{r}_1,\mathbf{r}_2;E) = \int &\frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}|C(\mathbf{q})|^2 \mathrm{e}^{i\mathbf{q}.(\mathbf{r}_1-\mathbf{r}_2)} \\ &\{(N_q+1)G^>(\mathbf{r}_1,\mathbf{r}_2;E-\hbar\omega_q) \\ &+N_qG^>(\mathbf{r}_1,\mathbf{r}_2;E+\hbar\omega_q)\} \end{aligned} \tag{28b}$$

where $C(\mathbf{q})$ describes the phonon wave-vector dependence of the matrix elements of the electron–phonon interaction and N_q denotes the Bose–Einstein distribution of the phonons assumed to remain in thermal equilibrium.

Using (28) in (21) and combining with (22) we construct

$$\begin{aligned} \Gamma^R(\mathbf{r}_1,\mathbf{r}_2;E) = \int &\frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}|C(\mathbf{q})|^2 \mathrm{e}^{i\mathbf{q}.(\mathbf{r}_1-\mathbf{r}_2)} \\ &\{(N_q+1)\pi A(\mathbf{r}_1,\mathbf{r}_2;E-\hbar\omega_q) \\ &+N_q\pi A(\mathbf{r}_1,\mathbf{r}_2;E+\hbar\omega_q)\} \end{aligned} \tag{29}$$

The real part of the retarded self-energy is then constructed from the dispersion relation (23).

The local approximation for the self-energy

For silicon, both interactions with acoustic phonons and intervalley optical phonons may be described accurately by local self-energies. For example, with elastic deformation potential acoustic phonon scattering at high temperatures

$$|C(\mathbf{q})|^2 = \frac{\hbar\Xi^2 q}{2\rho u_l} \text{ and } N_q \sim N_q + 1 \sim \frac{k_BT}{\hbar u_l q} \tag{30}$$

where Ξ is the deformation potential, ρ is the semiconductor density, u_l is the speed of sound. Expressions (23), (28), (29) for the self-energies then become *local* (diagonal in position space). In particular,

$$\begin{aligned} \Gamma^R(\mathbf{r}_1,\mathbf{r}_2;E) &= 2\pi D_{ac}^2 A(\mathbf{r}_1;E)\delta(\mathbf{r}_1-\mathbf{r}_2); \\ D_{ac}^2 &= \frac{\Xi^2 k_BT}{2\rho u_l^2} \end{aligned} \tag{31}$$

For non-polar optical phonon scattering, $|C(\mathbf{q})|^2$ is replaced by a constant C_0^2 and if we neglect the wave-vector

dependence of the optical phonon energy, the self-energies again become local, and in particular:

$$\Gamma^R(\mathbf{r}_1, \mathbf{r}_2; E) = \pi C_0^2\{(N_0+1)A(\mathbf{r}_1; E-\hbar\omega_0) + N_0 A(\mathbf{r}_1; E+\hbar\omega_0)\}\delta(\mathbf{r}_1-\mathbf{r}_2) \quad (32)$$

Since the spectral function A depends on G^R and hence the full complex retarded self-energy, it follows that Eqs. (23) and (32) have to be solved self-consistently.

Typically, the self-energies are initialized to zero and the SCBA equations are iterated until convergence, usually $\sim$ 100 iterations. This is compute-intensive as the self-energies are complex quantities.

However, the calculation of the *real part* of the self–energy is often ignored [6]. Perhaps, not surprisingly, as its calculation requires the evaluation of a Hilbert transform and it is very time-consuming. But, as we have seen, neglecting the real part of the self-energy violates the physical principle of antecedence (causality); as a consequence, the spectral density sum rule is violated and distortion of the density of states ensues [13,32,35]. In practice, the real part of the self-energy may be found [13] relatively efficiently with a discrete fast Fourier transform using a property that connects both Hilbert and Fourier transforms.

A simple Einstein model for the self-energy

An interesting model self-energy may be constructed from (32) by just considering phonon emission events. In the first Born approximation

$$\Gamma_1(E) = C_1 N_0(E-\hbar\omega)\theta(E-\hbar\omega) \quad (33)$$

For a 1D system, the density of states for a non-interacting system behaves as

$$N_0(E) = \frac{C_2}{\sqrt{E}}\theta(E)$$

$$\text{and thus } \Gamma_1 = \frac{C_3}{\sqrt{E-\hbar\omega}}\theta(E-h\omega) \quad (34)$$

where C_1, C_2, C_3 are constants. Using the dispersion relation we may compute

$$\Delta_1(E) = \frac{P}{\pi}\int_{-\infty}^{\infty} dE' \frac{\Gamma_1(E')}{E-E'} = C_3\frac{P}{\pi}\int_{\hbar\omega}^{\infty}\frac{dE'}{E-E'}\frac{1}{\sqrt{E'-\hbar\omega}} \quad (35)$$

Figure 22 shows the real and imaginary parts of the self-energy computed according to (34) and (35) for a relatively large coupling constant C_3.

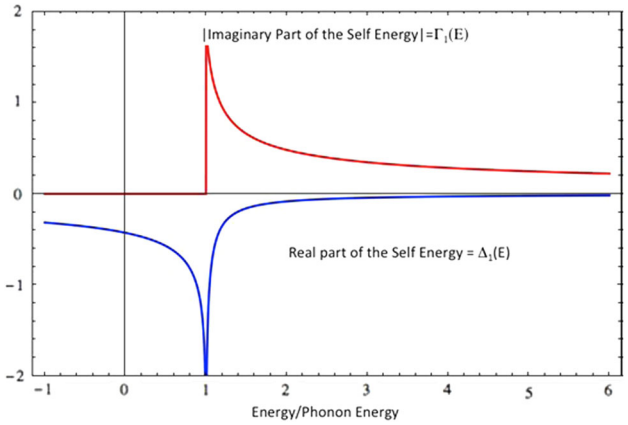

Fig. 22 Causal self-energy components in the Born approximation

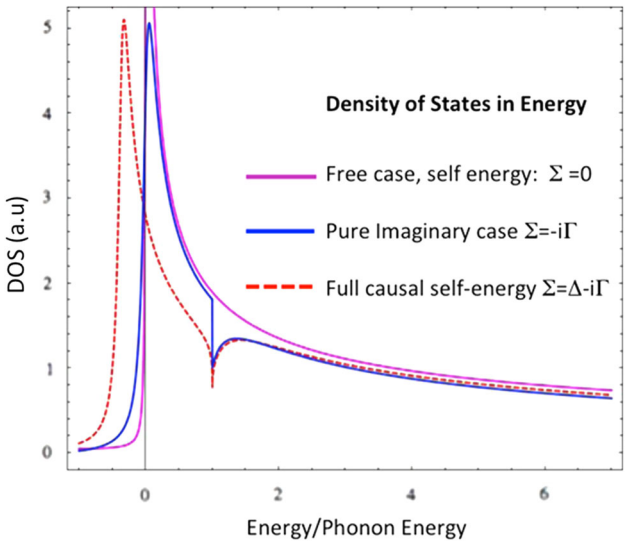

Fig. 23 Density of states in energy computed from (35) and (36)

Note that the real part extends across the whole spectrum and is by no means negligible. In Fig. 23 the resulting self-energy values are used to compute the density of states in energy with and without the real part of the self-energy and compared with the free density of states (34).

For this model the self-energy is a function of E. The spectral density function is most easily evaluated in the momentum representation giving:

$$A(p; E) = -\frac{1}{\pi}Im[E - E_p - \Delta^R(E) + i\Gamma^R(E) + i\eta]^{-1} = \frac{1}{\pi}\frac{\Gamma^R(E)+\eta}{[E-E_p-\Delta^R(E)]^2 + [\Gamma^R(E)+\eta]^2} \quad (36)$$

We keep the limiting causality factor η in (36) as this preserves causality even for regimes where $\Gamma^R = 0$ [32].

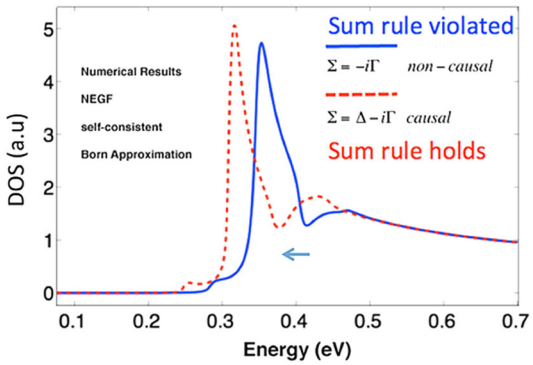

Fig. 24 Density of states in energy for simple model using full NEGF

The integral over energy is denoted N_S:

$$N_S = \int_{-\infty}^{\infty} A(p;E)\mathrm{d}E \tag{37}$$

It is easily verified by numerical computation that the spectral sum rule $N_S = 1$ is satisfied for the full self-energy (in the Born approximation) but the sum rule fails ($N_S < 1$) if we use the *acausal* approximation (real part of self-energy set to zero).

In obtaining these results the local density of states for a homogeneous infinite 1D system characterized by effective mass m is derived from the retarded Green function as:

$$G^R(x,x;E) = -\frac{1}{i\pi}\frac{m}{\hbar^2}\left\{\frac{\exp[ik(E)|x-x'|]}{k(E)}\right\} \tag{38}$$

$$k(E) = \sqrt{\frac{2m}{\hbar^2}}\sqrt{E-\Delta_1^R(E)+i\Gamma_1^R(E)+i\eta} \tag{39}$$

$$N_1(E) = -\frac{1}{\pi}Im[G^R(x,x;E)] = \frac{m}{\pi\hbar^2}\,\mathrm{Re}\left[\frac{1}{k(E)}\right]$$
$$= \sqrt{\frac{m}{2\pi^2\hbar^2}}\mathrm{Re}\left[\frac{1}{\sqrt{E-\Delta_1^R(E)+i\Gamma_1^R(E)+i\eta}}\right] \tag{40}$$

Evidently, if $\Sigma_1 \to 0$, $N_1(E) \to N_0(E) \sim \theta(E)/\sqrt{E}$.

The effect of adding the correct causal real part of the self-energy is to distort and shift the peak in the local density of states to lower energies while pinning the density of states at the phonon emission threshold. Qualitatively similar results are obtained by higher-order approximations as the SCBA is obtained by iteration. For comparison, the NEGF method was used with the same electron–phonon model for a long sample and the results shown in Fig. 24 are qualitatively very similar except for a net shift (including the phonon emission "singularity") to lower energies.

Non-locality of the self-energy

A more rigorous approach to phonon scattering reveals that even for elastic scattering and non-polar optical phonon scattering the self-energy is non-local on a scale of a few atoms wide. This has implications for ultra-scaled silicon devices. For polar optical phonon scattering in III–V materials the non-locality extends over a screening length. To illustrate we examine the contribution to the scattering kernel [48]. Along the transport direction x, extracted from (29):

$$K(x_1,x_2) = \int \frac{\mathrm{d}q_x}{(2\pi)^3}|C(q)|^2\,\mathrm{e}^{iq(x_1-x_2)} \tag{41}$$

The integration limits are determined by the first Brillouin zone as $\pm\pi/a$ where a is the lattice constant. If a were to be zero, the limits would be infinite and for $C(q)$ independent of q_x the kernel K would be localized:

$$K(x_1,x_2) = \frac{C^2}{4\pi^2}\int_{-\pi/a}^{\pi/a}\frac{\mathrm{d}q_x}{2\pi}\mathrm{e}^{iq_x(x_1-x_2)}$$
$$= \frac{C^2}{4\pi^2}\frac{\sin[\pi(x_1-x_2)/a]}{(x_1-x_2)} \tag{42}$$

$$K(x_1.x_2)(\lim a \to 0) \to C^2/(4\pi^2)\delta(x_1-x_2) \tag{43}$$

However, the *sinc* expression (42) shows that the *non-locality* extends to a few atomic spacings. In III–V materials, for screened polar optical phonon interactions, the scattering kernel may extend over distances up to the screening length. The net effect of using the local approximation for the scattering kernel is to underestimate the scattering rate as a function of energy [37,47].

Remote phonon scattering

Interestingly, an example of non-local interactions in silicon GAA structures is the remote phonon scattering mechanism which penetrates the channel via interaction of channel electrons with SO polar mode phonons at the dielectric–channel interface [48].

References

1. Yan, H., Choe, H.S., Nam, S.W., Hu, Y., Das, S., Klemic, J.F., Ellenbogen, J.C., Lieber, C.M.: Programmable nanowire circuits for nanoprocessors. Nature **470**, 240 (2011)
2. Suk, S.D., Li, M., Yeoh, Y.Y., Yeo, K.H., Cho, K.H., Ku, I.K., Cho, H., Jang, W.J., Kim, D.-W., Park, D., Lee, W.-S.: Investigation of nanowire size dependency on TSNWFET. In: 2007 IEDM Tech. Dig., pp. 891–894 (2007)

3. Ieong, M., Doris, B., Kedzierski, J., Rim, K., Yang, M.: Science **306**, 2057 (2004)
4. Aldegunde, M., Martinez, A., Asenov, A.: Non-equilibrium Green's function analysis of cross section and channel length dependence of phonon scattering and its impact on the performance of Si nanowire field effect transistors. J. Appl. Phys. **110**(9), 094518 (2011)
5. Kotlyar, R., Obradovic, B., Matagne, P., Stettler, M., Giles, M.D.: Assessment of room-temperature phonon-limited mobility in gated silicon nanowires. Appl. Phys. Lett. **84**(25), 5270 (2004)
6. Jin, S., Park, Y.J., Min, H.S.: A three-dimensional simulation of quantum transport in silicon nanowire transistor in the presence of electron-phonon interactions. J. Appl. Phys. **99**(12), 123719 (2006)
7. Jin, S., Fischetti, M.V., Tang, T.-W.: Modeling of electron mobility in gated silicon nanowires at room temperature: surface roughness scattering, dielectric screening, and band nonparabolicity. J. Appl. Phys. **102**, 083715 (2007)
8. Buin, A., Verma, A., Anantram", M.P.: Carrier–phonon interaction in small cross-sectional silicon nanowires. J. Appl. Phys. **104**, 053716 (2008)
9. Ke, M.: A Study of Carrier Mobility and Variability in Silicon Nanowire MOSFETs. PhD diss. (2012)
10. Esposito, A., Frey, M., Schenk, A.: Quantum transport including nonparabolicity and phonon scattering: application to silicon nanowires. J. Comput. Electron. **8**, 336 (2009)
11. Svizhenko, A., Anantram, M.: Effect of scattering and contacts on current and electrostatics in carbon nanotubes. Phys. Rev. B **72**, 085430 (2005)
12. Aldegunde, M., et al.: Dependence of Matthiessen's rule on complex phonon self-energies: a NEGF study. In: Proceeding of 16th International Workshop on Computational Electronics, June 4–7, p. 220. Nara, Japan (2013)
13. Valin, R., Aldegunde, M., Martinez, A., Barker, J.R.: Quantum transport of a nanowire field-effect transistor with complex phonon self-energy. J. Appl. Phys. **116**(8), 084507 (2014)
14. King, F.W.: Hilbert Transforms, vol. 1. Cambridge University Press, Cambridge (2009)
15. Zimbovskaya, N.A.: Transport properties of molecular junctions. In: Springer Tracts in Modern Physics, vol. 254, pp. 46. Springer-Verlag, New York (2013)
16. Kronig, R.L.: On the theory of the dispersion of X-rays. J. Opt. Soc. Am. **12**, 547–557 (1926). doi:10.1364/JOSA.12.000547
17. Kramers, H.A.: La diffusion de la lumiere par les atomes. Atti Cong. Intern. Fisici (Transactions of Volta Centenary Congress) Como. **2**, 545–557 (1927)
18. Swaminathan, K., Kotra, J., Liu, H., Sampson, J., Kandemir, M., Narayanan, V.: Thermal-aware application scheduling on device-heterogeneous embedded architectures. In: 2015 28th International Conference on VLSI Design (VLSID), pp. 221–226. IEEE (2015)
19. Lundstrom, M.: Moore's law forever? Science **299**(5604), 210 (2003)
20. Moore, A.L., Shi, L.: Emerging challenges and materials for thermal management of electronics. Mater. Today **17**(4), 163–174 (2014)
21. Lake, R., Datta, S.: Nonequilibrium Green's-function method applied to double-barrier resonant-tunneling diodes. Phys. Rev. B **45**(12), 6670 (1992)
22. Lake, R., Datta, S.: Energy balance and heat exchange in mesoscopic systems. Phys. Rev. B **46**(8), 4757–4763 (1992)
23. Mahan, G.: Quantum transport equation for electric and magnetic fields. Phys. Rep. **145**, 251 (1987)
24. Martinez, A., Barker, J.R., Aldegunde, M., Valin, R.: Study of local power dissipation in ultrascaled silicon nanowire FETs. IEEE Electron Device Lett. **36**(1), 2–4 (2015)
25. Joyce, H.J., Gao, Q., Tan, H.H., Jagadish, C., Kim, Y., Zou, J., Smith, L.M., Jackson, H.E., Yarrison-Rice, J.M., Parkinson, P., Johnston, M.B.: III-V semiconductor nanowires for optoelectronic device applications. Prog. Quantum Electron. **35**(2), 23–75 (2011)
26. Svizhenko, A., Anantram, M.P., Govindan, T.R., Biegel, B., Venugopal, R.: Two-dimensional quantum mechanical modeling of nanotransistors. J. Appl. Phys. **91**, 2343 (2002)
27. Martinez, A., Bescond, M., Barker, J., Svizhenko, A., Anantram, M., Millar, C., Asenov, A., Trans, I.E.E.E.: A self-consistent full 3-D real space NEGF simulator for studying nonperturbative effects in nano-Mosfets. Electron Dev. **54**, 2213 (2007)
28. Poli, S., Pala, M.: Channel length dependence of low field mobility in silicon nanowires FETs. IEEE Electron Dev. Lett. **30**, 1212 (2009)
29. Lenzi, M., Palestri, P., Gnani, E., Reggiani, S., Gnudi, A., Esseni, D., Selmi, L., Baccarani, G., Trans, I.E.E.E.: Investigation of the transport properties of silicon nanowires using deterministic and Monte Carlo approaches to the solution of the Boltzmann transport equation. Electron Dev. **55**, 2086 (2008)
30. Luisier, M., Klimeck, G.: Atomistic full-band simulations of silicon nanowire transistors: effects of electron-phonon scattering. Phys. Rev. B **80**(15), 155430 (2009)
31. Niquet, Y.M., Delerue, C., Rideau, D., Videa, B.: Fully atomistic simulations of phonon-limited mobility of electrons and holes in $< 001 > -$, $< 110 > -$, and $< 111 > -$oriented Si nanowires. IEEE Trans. Electron Dev. **59**(5), 1480–1487 (2012)
32. Barker, J.R., Martinez, A.: Self-energy models for scattering in semiconductor nanoscale devices: causality considerations and the spectral sum rule? In: Material Research Society Proceedings, MRS Online Proceedings Library, 1551, mrss13-1551-r10-25. doi:10.1557/opl.2013.898 (2013)
33. Singha, P.: Deo, nondispersive backscattering in quantum wires. Phys. Rev. B **75**, 235330 (2007)
34. Friedel, J.: The distribution of electrons round impurities in monovalent metals. Philos. Mag. **43**, 153 (1952)
35. Barker, J.R., Martinez, A., Aldegunde, M., Valin, R.: Causal self-energies for NEGF modelling of quantum nanowires. J. Phys. Conf. Ser. **526**, 012001 (2014)
36. Rhyner, R., Luisier, M.: Influence of anharmonic phonon decay on self-heating in Si nanowire transistors. Appl. Phys. Lett. **105**, 062113 (2014)
37. Smith, A.C., Janak, J.F., Adler, R.B.: Electronic Conduction Solids. McGraw-Hill, New York (1967)
38. Niquet, Y.M., Lherbier, A., Quang, N.H., Fernández-Serra, M.V., Blase, X., Delerue, C.: Electronic structure of semiconductor nanowires. Phys. Rev. B **73**, 165319 (2006)
39. Kubis, T.: Quantum transport in semiconductor nanostructures. In: Abstreiter, G., Amann, M.C., Stutzmann, M., Vogl, P. (eds.) Verein zur Foerderung des Walter Schottky Instituts der Technischen Universitaet Muenchen e.V., Garching (2009)
40. Lundstrom, M.: Fundamentals of Carrier Transport. Cambridge University Press, Cambridge (2000)
41. Price, A., Martinez, A.: Investigation on phonon scattering in a GaAs nanowire field effect transistor using the non-equilibrium Green's function formalism. J. Appl. Phys. **164501**, 164501 (2015)
42. Price, A., Martinez, A., Valin, R., Barker., J.R.: Impact of different electron-phonon scattering models on the electron transport in a quantum wire. J. Phys. Conf. Ser. **526**(1), 012007. IOP Publishing (2014)
43. Mera, H., Lannoo, M., Li, C., Cavassilas, N., Bescond, M.: Inelastic scattering in nanoscale devices: one-shot current-conserving lowest-order approximation. Phys. Rev. B. **86**(16), 161404 (2012)
44. Bescond, M.: Modeling of phonon scattering in n-type nanowire transistors using one-shot analytic continuation technique. J. Appl. Phys. **114**(15), 153712 (2013)
45. Datta, S.: Electronic Transport In Mesoscopic Systems. Cambridge University Press, New York (1995)

46. Kadanoff, L.P., Baym, G.: Quantum Statistical Mechanics. Benjamin, New York (1962)
47. Barker, J.R., Martinez, A., Aldegunde, M.: Compliant energy and momentum conservation in NEGF simulation of electron–phonon scattering in semiconductor nano-wire transistors. J. Phys. Conf. Ser. **367**, 012012 (2012)
48. Barker, J.R., Martinez, A.: Remote soft-optical phonon scattering in Si nanowire FETs. In: IEEE Proceedings of International Workshop on Computational Electronics 2014, Curran Associates (ISBN:9781479954346), pp. 143–145 (2014). Also on-line in IEEE Xplore doi:10.1109/IWCE.2014.6865851